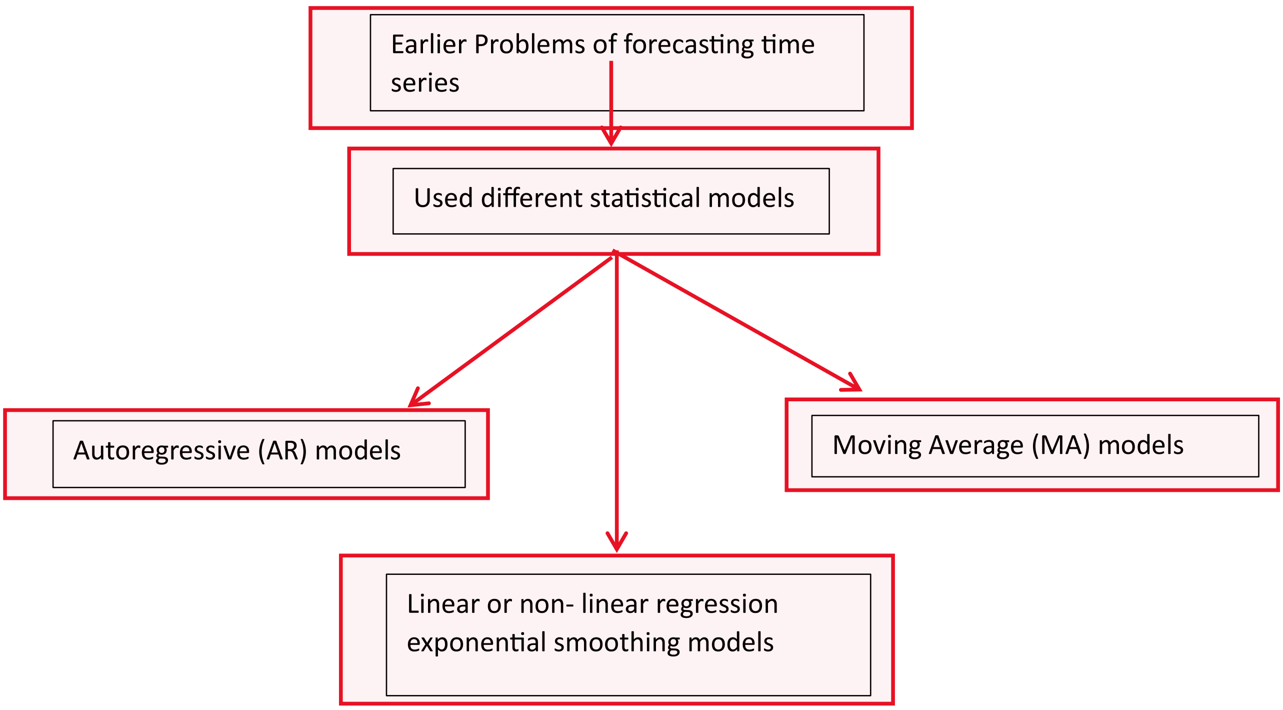
Name:- Aryan Gupta

Branch:- CSE- 22074007

PROBLEM :- TO FIND A NOVEL ACTIVATION FUNCTION FOR TEMPORAL DEPENDENCIES

- Artificial neural networks (ANN) -> This is used in studies of complex time series forecasting like weather, energy consumption and financial series.
- Features of a time series in finance & economics :-
    - Data Intensity
    - Unstructured nature
    - High degree of uncertainity
    - Hidden relationships
- Performance of Neural Networks depends on :-
    - No. of hidden layers
    - No. of hidden nodes
    - Learning algorithm
    - Activation Functions



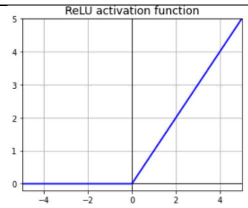
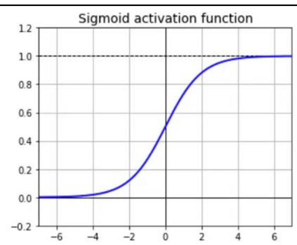
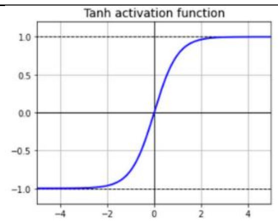
Need for non-linearity in neural networks:

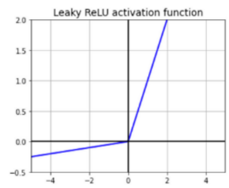
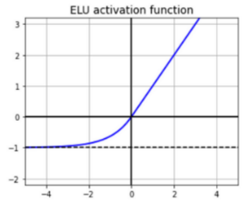
- we need to apply an activation function to make the network dynamic and add the ability to it to extract complex and complicated information from data
- represent non linear convoluted random functional mappings between input and output.
- An important feature of an activation function is that it must be differentiable so that we can implement back propagation optimization strategy in order to compute the errors or losses with respect to weights and eventually optimize weights using Gradient Descend or any other optimization technique to reduce errors

Problems with activation functions

- Saturation
- Vanishing gradient descent

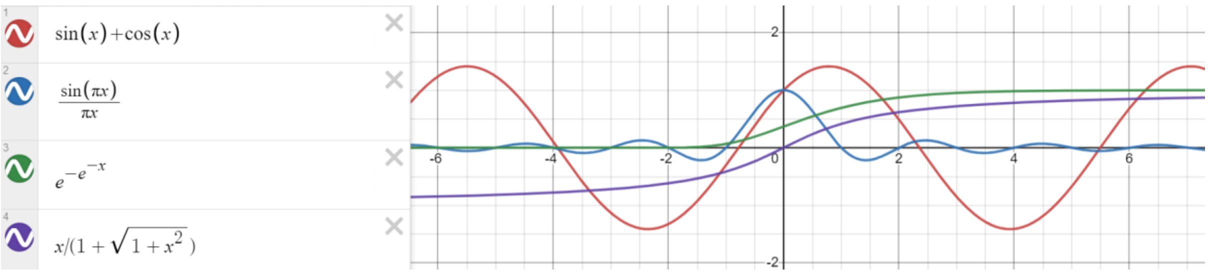
Some commonly used activation functions :-

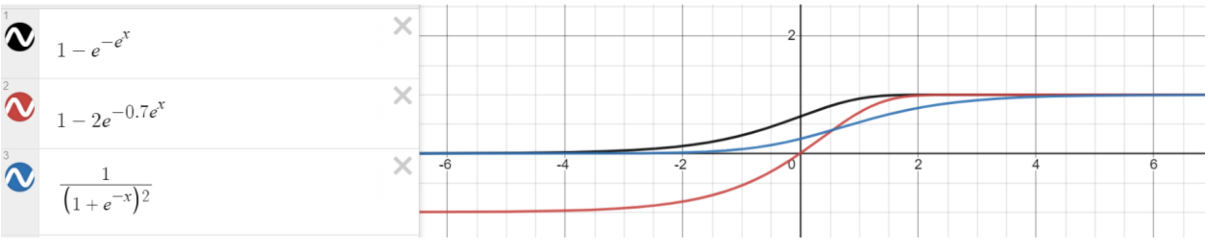
RELU	$\begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$ 
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$ 
Hyperbolic Tangent(Tanh)	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ 

Leaky RELU	f(x) = max(αx, x)	
Exponential Linear Unit (ELU)	$\begin{cases} \alpha\left(e^{x}-1\right) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$	

Some Complex Activation Functions

sinc	$\frac{sin(\pi x)}{\pi x}$
sincos	sin(x) + cos(x)
root sig	$\frac{x}{1+\sqrt{1+x^2}}$
Log sigm	$\frac{1}{(1+e^{-x})^2}$
Log log	$e^{-e^{-x}}$
Clog log	$1-e^{-e^{x}}$
Log logm	$1-2e^{-0.7e^{x}}$





Error function used in the research papers –

Mean average percentage error (MAPE%)

MAPE % = $\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100$

Where A_t is the actual value and F_t is the forecast value.

Algorithms with best performing activation functions:-

- CGF Algorithm :-
  - cloglog,cloglogm and loglog with smaller networks {2,4,6}
  - cloglog,cloglogm improved performance when data was asymmetric
  - log sig,tanh,rootsig better performance with larger networks {8,12,16,20}
- LM Algorithm :-
  - logsig activation function
  - LM is more expensive than CGF

The dataset used in the project is taken from Kaggle datasets :- https://www.kaggle.com/datasets/atechnohazard/battery-and-heating-data-in-real-driving-cycles

About Dataset :-

High-voltage batteries in battery electric vehicles face significant load fluctuations due to driving behavior. This dynamic performance of the powertrain is contrasted by the almost constant load of the auxiliary consumers. The highest auxiliary consumption is generated by the heating and air conditioning system, which decreases the vehicles range significantly. 72 real driving trips with a BMW i3 (60 Ah) were recorded, serving for model validation of a full vehicle model consisting of the powertrain and the heating circuit.

Each trip contains:

- Environmental data (temperature, elevation, etc.)

- Vehicle data (speed, throttle, etc.)
- Battery data (voltage, current, temperature, SoC)
- Heating circuit data (indoor temperature, heating power, etc.)

The measurement data is in CSV format to ensure easy use by various standard programs (Excel, Matlab) or own developed codes (Python).

The script "readin.m" reads the csv files into MatLab as table and struct.

The data contains vehicle data, battery data and heating data.

The measurement data is divided into two categories. Category A was recorded in summer and does not contain all measured data due to trouble with the measurement system. Category B was recorded in winter and is consistent.

For the sake of simplicity we took data of 12 trips due to computational limitations.

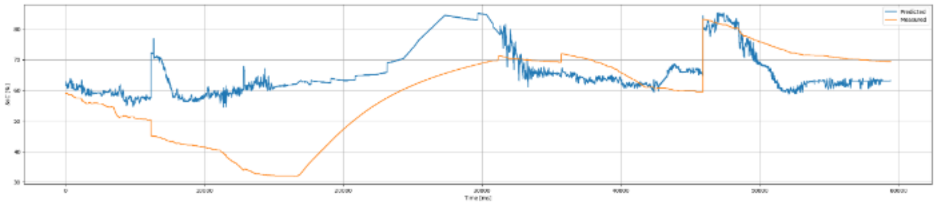
The data contains 198053 rows × 48 columns.

Application of Different Activation Functions on the Dataset

	Activation Functions	RMSE Loss	MAE Loss
1	Relu	13.784	11.131
2	LeakyRelu	16.631	12.013
3	Elu	19.125	14.299
4	Tanh	18.853	14.083
5	Sinc	18.815	14.186
6	Root sig	19.325	14.314
7	Min(x,sinx)	16.374	11.653
8	x/(1+x^2)	12.426	10.32
9	x/(1+x^2)^(0.5)	19.273	14.592
10	x/(1+x)	24.48	18.209

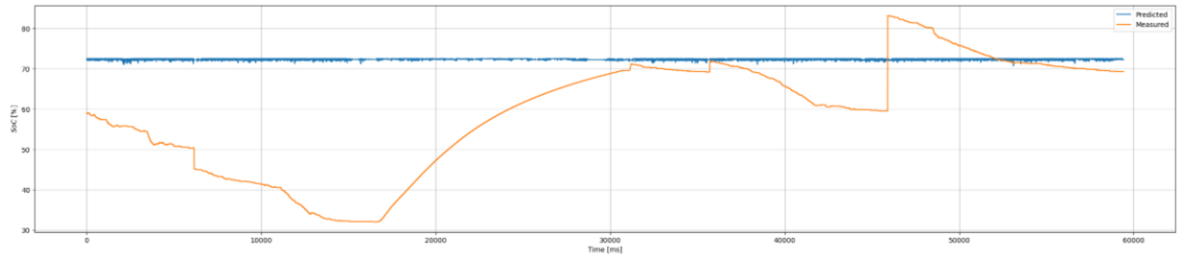
1. Relu

```
Neural Network:
Root Mean Square Error: 13.784660493308435
Mean Absolute Error: 11.13097748774257
```



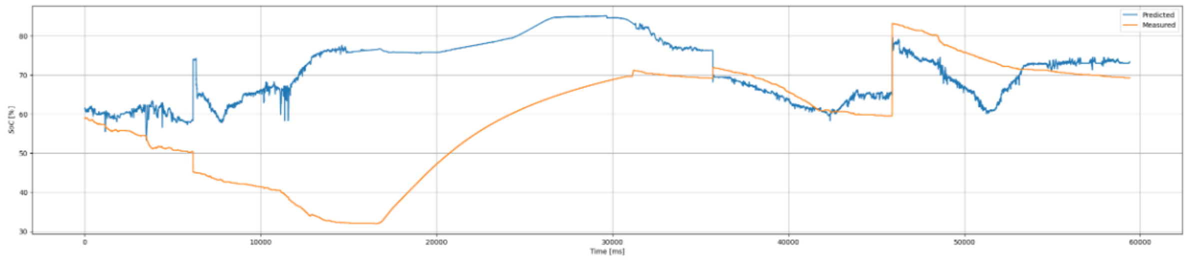
2. LeakyRelu

```
Neural Network:
Root Mean Square Error: 18.94195296565083
Mean Absolute Error: 14.174227586085951
```



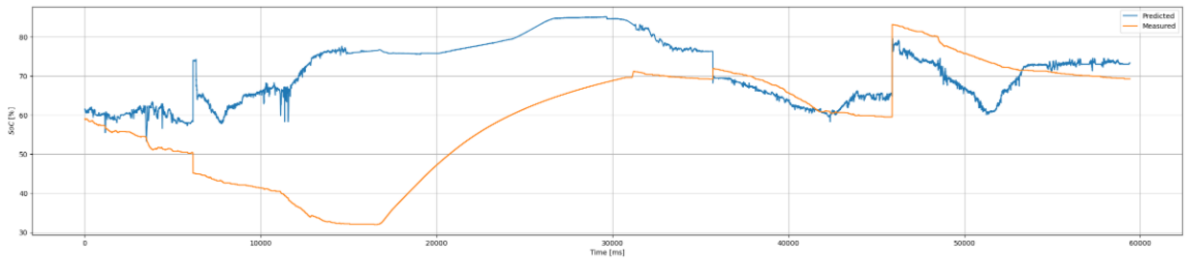
3. Elu

```
Neural Network:
Root Mean Square Error: 19.125255789253984
Mean Absolute Error: 14.299106409235439
```



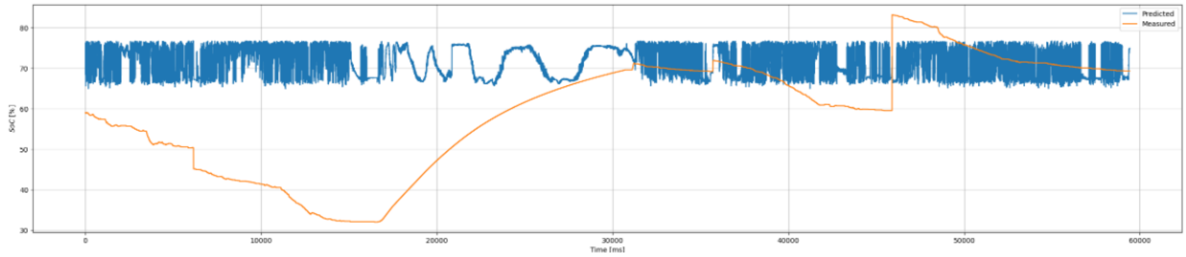
4. Min(x,sinx)

```
Neural Network:
Root Mean Square Error: 19.125255789253984
Mean Absolute Error: 14.299106409235439
```



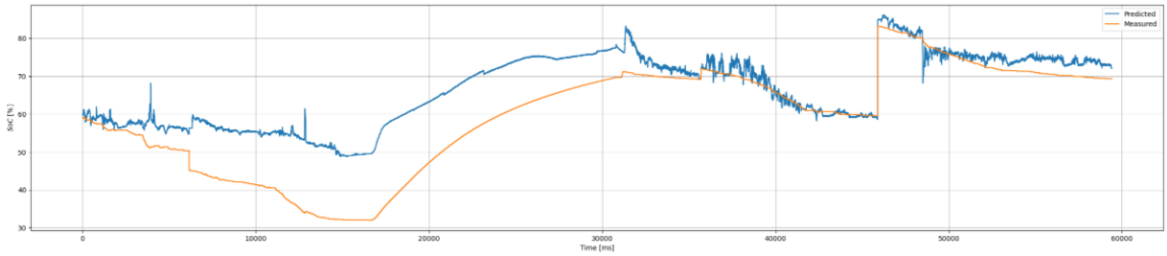
5. Sinc

```
Neural Network:
Root Mean Square Error: 18.81516484709987
Mean Absolute Error: 14.186075246309354
```



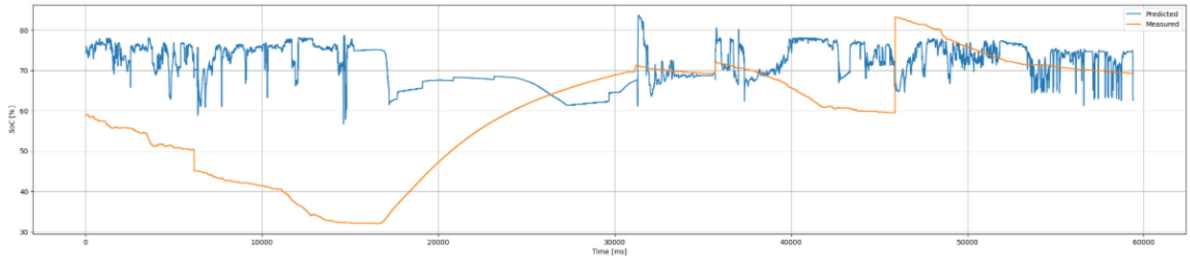
6. $\frac{x}{(1+x^2)}$

```
Neural Network:
Root Mean Square Error: 9.862316848102665
Mean Absolute Error: 7.43838905893811
```



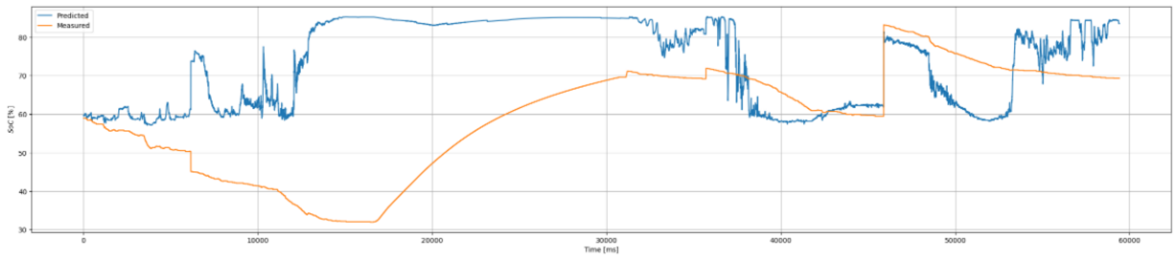
7. Root sig

```
1857/1857 [==============================] - 3s 1ms/step
Neural Network:
Root Mean Square Error: 19.325547743381918
Mean Absolute Error: 14.314459909911879
```



8. $\frac{x}{1+\sqrt{1+x^2}}$

```
1857/1857 [==============================] - 2s 1ms/step
Neural Network:
Root Mean Square Error: 22.54696490268259
Mean Absolute Error: 17.376286485029194
```



From the given graphs and result we can analyze that the activation functions min(x,sinx) and $\frac{x}{(1+x^2)}$ when parameterized can be used as a novel activation function for the dataset.

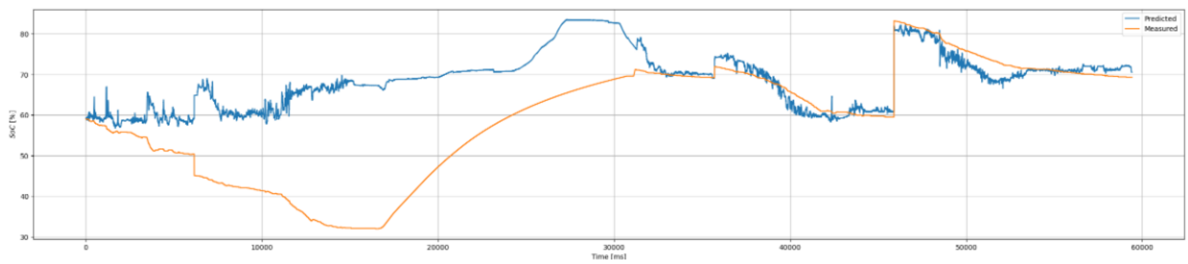
The activation function clog log, log log, log sig are not working on the given dataset giving Nan or infinity error.

Parameterization of the Activation Function :- min(x,sinx)

	Activation Function Min(x,sinx)		
1	min(2x,sinx)	15.155	10.509
2	min(x/2,sinx)	13.784	11.131
3	min(x/2,2sinx)	15.909	11.756
4	min(x/2,sinx/2)	16.668	12.232

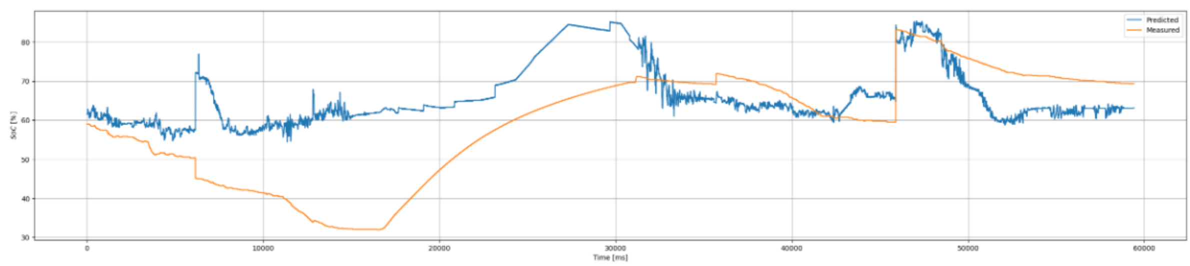
a) Min(2x,sinx)

```
Neural Network:
Root Mean Square Error: 15.155234593722847
Mean Absolute Error: 10.509974946463366
```



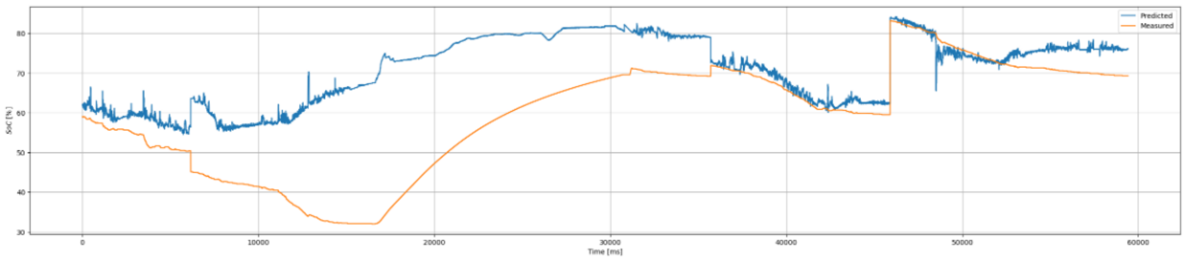
b) Min(x/2,sinx)

```
Neural Network:
Root Mean Square Error: 13.784660493308435
Mean Absolute Error: 11.13097748774257
```



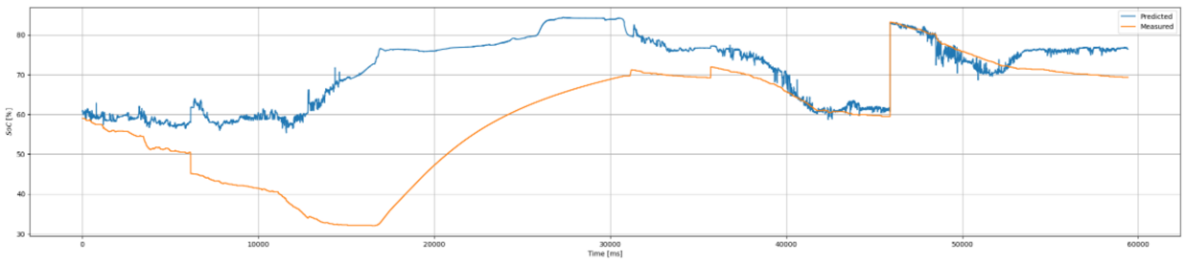
c) Min(x/2,2sinx)

```
Neural Network:
Root Mean Square Error: 15.909883813904308
Mean Absolute Error: 11.756552170719708
```



d) Min(x/2,sinx/2)

```
Neural Network:
Root Mean Square Error: 16.668266805287907
Mean Absolute Error: 12.232081429784673
```



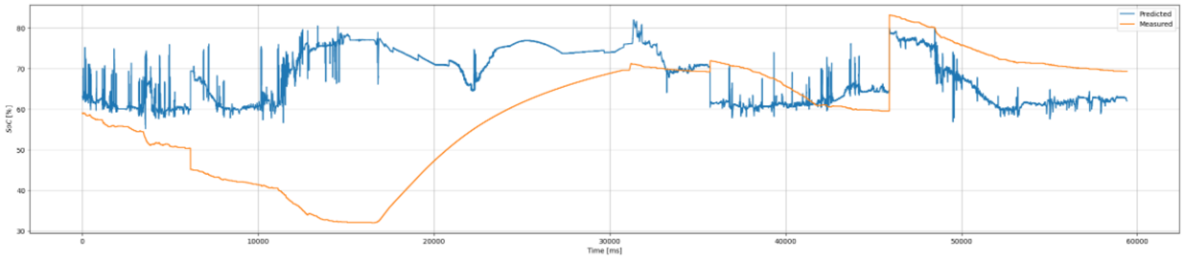
Parameterization of the Activation Function :-

$$\frac{x}{(1+x^2)}$$

	Activation Function x/(1+x^2)		
1	x/(1+2*x^2)	17.904	13.439
2	x/(1+(x^2)/2)	8.715	6.527
3	2x/((1+x*x/2))	11.26	8.759
4	x/(1+x*x/1.5)	9.485	7.194
5	3x/(1+x*x/1.5)	12.386	10.354
6	1.5x/(1+x*x/1.5)	20.202	14.715

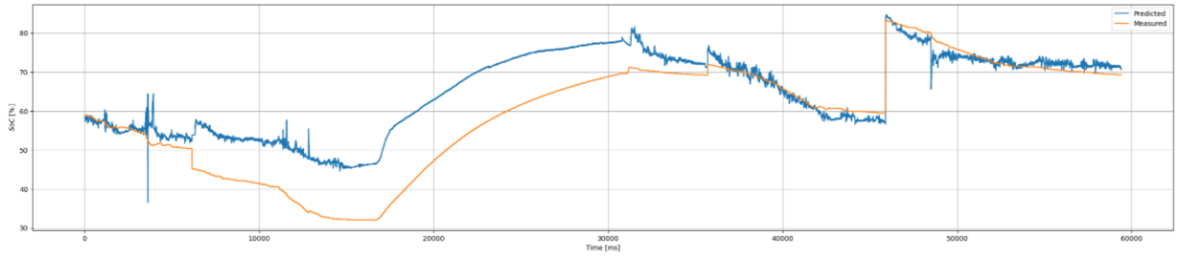
a) $\frac{x}{(1+2x^2)}$

```
Neural Network:
Root Mean Square Error: 17.904183044285627
Mean Absolute Error: 13.439212654270937
```



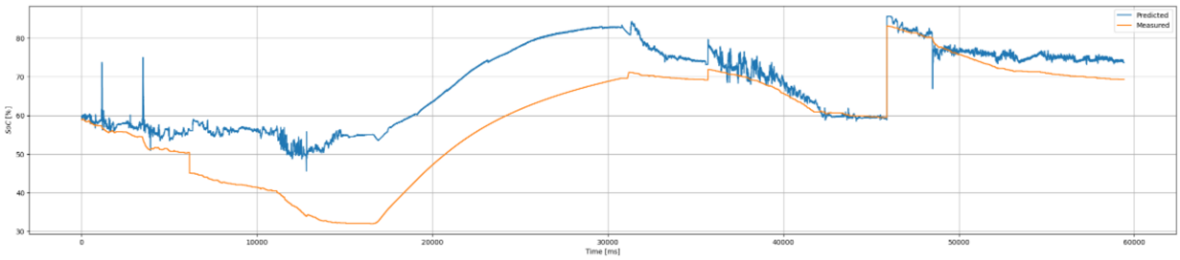
b) $\frac{x}{(1+(x^2)/2)}$

```
Neural Network:
Root Mean Square Error: 8.715673578040978
Mean Absolute Error: 6.5279872794669425
```



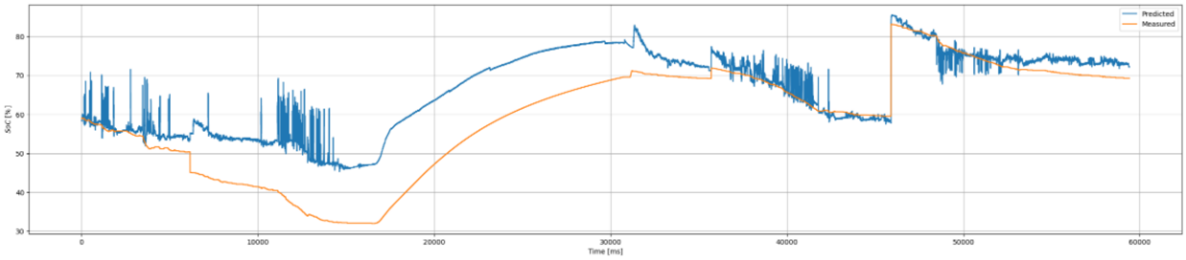
c) $\frac{2x}{(1+(x^2)/2)}$

```
Neural Network:
Root Mean Square Error: 11.260657211459248
Mean Absolute Error: 8.759808102082932
```



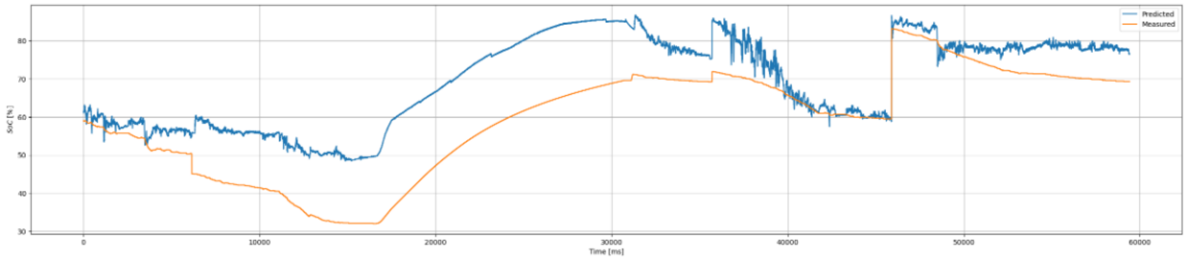
d) $\frac{(x)}{(1+(x^2)/1.5)}$

Neural Network:
Root Mean Square Error: 9.485457653788433
Mean Absolute Error: 7.1945222085716765



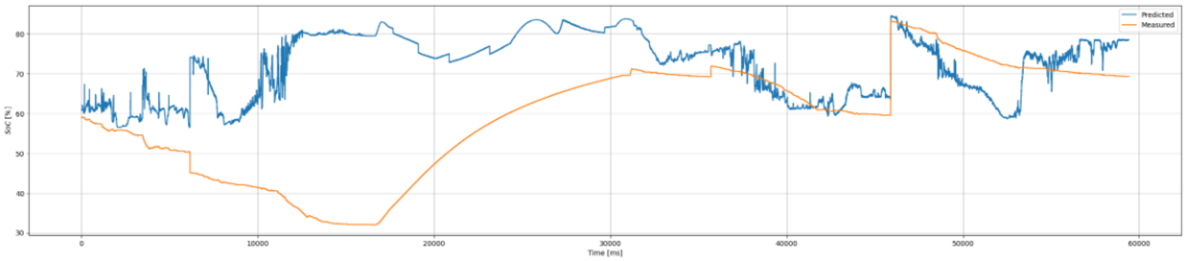
$e)\ \frac{(3x)}{(1+(x^2)/1.5)}$

Neural Network:
Root Mean Square Error: 12.386274568801174
Mean Absolute Error: 10.35452185025362



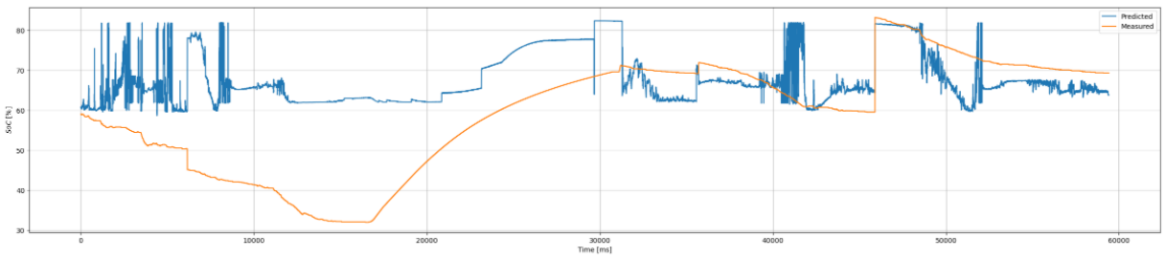
$f)\ \frac{(1.5x)}{(1+(x^2)/1.5)}$

Neural Network:
Root Mean Square Error: 20.20225596603635
Mean Absolute Error: 14.715039897128744



g) $\frac{x}{(1+(x^2)/2)} + 1$

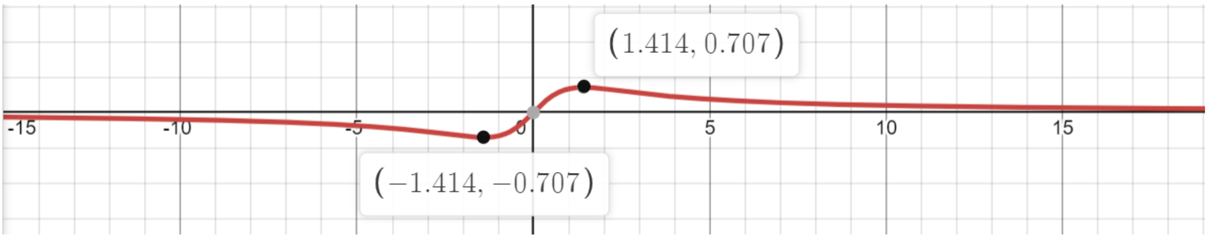
Neural Network:
Root Mean Square Error: 14.93472066244417
Mean Absolute Error: 11.623239399884419



As a result of the given analysis we can conclude that $\frac{x}{(1+(x^2)/2)}$ is the most appropriate activation function to be used as it has least Root Mean Square Error and Mean Absolute Error.

Differentiability :- $\frac{2(2-x^2)}{(2+x^2)^2}$

The function is differentiable at x=0.



Sources :-

- https://www.ijeast.com/papers/310-316,Tesma412,IJEAST.pdf
- https://www.sciencedirect.com/science/article/pii/S1319157820303505
- https://link.springer.com/article/10.1007/s00521-010-0407-3
- https://www.sciencedirect.com/science/article/pii/S0925231220318749
- https://www.sciencedirect.com/science/article/pii/S0893608021000344